

Frequency Coverage Properties of a Uniform Shrinkage Prior Distribution

Hyungsuk Tak*

Abstract. A uniform shrinkage prior (USP) distribution on the unknown variance component of a random-effects model is known to produce good frequency properties. The USP has a parameter that determines the shape of its density function, but it has been neglected whether the USP can maintain such good frequency properties regardless of the choice for the shape parameter. We investigate which choice for the shape parameter of the USP produces Bayesian interval estimates of random effects that meet their nominal confidence levels better than several existent choices in the literature. Using univariate and multivariate Gaussian hierarchical models, we empirically show that the USP can achieve its best frequency properties when its shape parameter makes the USP behave similarly to an improper flat prior distribution on the unknown variance component.

Keywords: multi-level model, random effects, overdispersion.

1 Introduction

We briefly review a uniform shrinkage prior (USP) using a multivariate Gaussian hierarchical model (or a multivariate Normal-Normal model) (Daniels, 1999; Everson and Morris, 2000). We use notation $\mathbf{Y}_j = (Y_{1j}, Y_{2j}, \dots, Y_{pj})^\top$ to denote a vector of p unbiased estimators of group j for a vector of p random effects $\boldsymbol{\theta}_j = (\theta_{1j}, \theta_{2j}, \dots, \theta_{pj})^\top$ with a two-level structure as follows. For $j = 1, 2, \dots, k$,

$$\mathbf{Y}_j \mid \boldsymbol{\theta}_j \stackrel{\text{indep.}}{\sim} N_p(\boldsymbol{\theta}_j, \mathbf{V}_j), \quad (1)$$

$$\boldsymbol{\theta}_j \mid \mathbf{A}, \boldsymbol{\beta} \stackrel{\text{indep.}}{\sim} N_p(\mathbf{X}_j^\top \boldsymbol{\beta}, \mathbf{A}), \quad (2)$$

where \mathbf{V}_j is a known $p \times p$ covariance matrix, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_{mp})^\top$ is a vector for unknown mp regression coefficients, \mathbf{X}_j is a known $mp \times p$ block diagonal matrix with $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jm})^\top$, a covariate vector of length m for group j , repeatedly appearing along the diagonal, and \mathbf{A} is an unknown $p \times p$ covariance matrix of the conjugate Gaussian prior distribution. The corresponding conditional posterior distribution for group j is

$$\boldsymbol{\theta}_j \mid \mathbf{A}, \boldsymbol{\beta}, \mathbf{y}_j \sim N_p((I_p - \mathbf{B}_j)\mathbf{y}_j + \mathbf{B}_j \mathbf{X}_j^\top \boldsymbol{\beta}, (I_p - \mathbf{B}_j)\mathbf{V}_j), \quad (3)$$

where \mathbf{y}_j is the observed unbiased estimate, I_p is a p -dimensional identity matrix, and $\mathbf{B}_j \equiv \mathbf{V}_j(\mathbf{V}_j + \mathbf{A})^{-1}$ is a positive definite $p \times p$ matrix for the shrinkage factor of group j . Daniels (1999) defines the multivariate USP as a Uniform distribution on the space of the positive definite shrinkage matrix $\mathbf{B}_0 = \mathbf{V}_0(\mathbf{V}_0 + \mathbf{A})^{-1}$ whose eigenvalues are

* Statistical and Applied Mathematical Sciences Institute, hyungsuk.tak@gmail.com

between $(0, 1]$; \mathbf{V}_0 is a known positive definite symmetric matrix. The density function of the USP with respect to \mathbf{A} that is transformed from the USP on \mathbf{B}_0 is proportional to

$$|\mathbf{V}_0 + \mathbf{A}|^{-p-1} \times I_{\{|\mathbf{A}| > 0\}} d\mathbf{A}, \quad (4)$$

where $I_{\{w\}}$ is an indicator function of w . The known matrix \mathbf{V}_0 determines the shape of this density function in (4) and thus we call it the shape parameter of the USP. (Hereafter, a USP is always with respect to \mathbf{A} .) An approximate USP in generalized linear mixed models (Natarajan and Kass, 2000), derived by replacing the first-level distributions for \mathbf{Y}_j with an MLE-based Gaussian approximation, is the same as (4) with a specific value for \mathbf{V}_0 . (See also Chen and Wehrly (2016) for the approximate USP in multivariate generalized linear mixed models.) For a univariate case ($p = 1$), we use unbolded notation, i.e., Y_j , y_j , θ_j , V_j , V_0 , A , B_j , and B_0 .

The USP is known to have several attractive properties. Strawderman (1971) showed using a Normal-Normal model ($p = 1$) that the unconditional posterior mean of a random effect, i.e., $E(\theta_j | y_j)$, based on the USP leads to an admissible minimax estimator if $k > 5$. Constructing a Poisson-Gamma hierarchical model ($p = 1$), Christiansen and Morris (1997) indicated that the USP is relatively non-informative with good frequency coverage properties in random effects estimation. In particular, they mentioned that the shape parameter of the USP could be adjusted to have more conservative results in estimating random effects, which we empirically prove in this article. Using a Normal-Normal model ($p = 1$), Daniels (1999) compared the frequency properties of a USP with those of several non-informative priors such as improper flat and Jeffreys' priors under the settings for hypothesis testing and estimation on A . Natarajan and Kass (2000) showed good frequency coverage properties of the approximate USP compared to other default priors, e.g., approximate Jeffreys' prior, using non-Gaussian data.

Many authors have used various choices for the shape parameter of the USP. Albert (1988) set $V_0 = 1$ for a hierarchical generalized linear model, although there was no specific reason for this choice; the article focused on an accurate approximation to posterior distributions of A and β . Christiansen and Morris (1997) suggested setting the shape parameter to a median value of the random-effects variance component for their Poisson-Gamma model as a less informative choice. DuMouchel (1994) suggested setting V_0 to a harmonic mean of $\{V_1, V_2, \dots, V_k\}$ as a typical value of V_j in a univariate Normal-Normal model. Following DuMouchel (1994), Daniels (1999) set \mathbf{V}_0 to the harmonic mean of $\{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_k\}$ in a multivariate Normal-Normal model. This harmonic mean was also adopted for the approximate USP (Natarajan and Kass, 2000). Besides the harmonic mean, Everson and Morris (2000) set \mathbf{V}_0 to an arithmetic mean of $\{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_k\}$ as another typical value of \mathbf{V}_j to mimic an equal variance case.

However, it has been neglected whether the USP maintains good frequency properties regardless of the choice for its shape parameter, and no one has investigated the frequency coverage properties of the USP in a multivariate case ($p \geq 2$). This article investigates the frequency coverage properties of the USP according to the shape parameter, using univariate and bivariate Normal-Normal models. To do this, we extend a univariate repeated sampling coverage evaluation (Tak et al., 2017) to a multivariate one to see which choice for the shape parameter produces Bayesian interval estimates for

random effects that meet their nominal confidence levels better. Our numerical illustrations show that the best frequency coverage properties of the USP can be achieved when the shape parameter lets the USP approach an improper flat prior distribution on \mathbf{A} . Figure 1 displays one numerical illustration in Section 4.1. It shows that a Normal-Normal model based on a USP with V_0 set to a harmonic mean of $\{V_1, V_2, \dots, V_k\}$ denoted by $V_{0,DM}$ (DuMouchel, 1994; Daniels, 1999; Natarajan and Kass, 2000) produces posterior intervals for random effects that do not meet the 95% confidence level. On the other hand, a choice for V_0 that flattens the USP on \mathbf{A} with $V_0 = 10^4 \times V_{0,DM}$ achieves the spirit of the 95% confidence level better. See Section 4.1 for details.

The rest of this article is organized as follows. We introduce a multivariate version of the frequency coverage evaluation in Section 3. In Section 4, we explain which choices for the shape parameter achieve the spirit of the confidence level better and why, using eight school data for a simple univariate case (Rubin, 1981) and twenty seven hospital profiling data for a bivariate case (Everson and Morris, 2000).

2 A Markov chain Monte Carlo method

Our target distribution is the joint posterior distribution of the unknown parameters, $\theta \equiv \{\theta_1, \theta_2, \dots, \theta_k\}$, \mathbf{A} , and β , whose probability density function is defined as

$$\pi(\theta, \mathbf{A}, \beta | \mathbf{y}) \propto h(\mathbf{A}, \beta) \times \prod_{j=1}^k [f(\mathbf{y}_j | \theta_j) \times g(\theta_j | \mathbf{A}, \beta)], \quad (5)$$

where $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k\}$, the distributions for $f(\mathbf{y}_j | \theta_j)$ and $g(\theta_j | \mathbf{A}, \beta)$ are specified in (1) and (2), respectively, and the joint prior density function $h(\mathbf{A}, \beta)$ is proportional to $|\mathbf{V}_0 + \mathbf{A}|^{-p-1} \times I_{\{|\mathbf{A}| > 0\}} d\mathbf{A} d\beta$, meaning that the prior distribution on \mathbf{A} is the USP and independently the prior distribution on β is an improper flat distribution

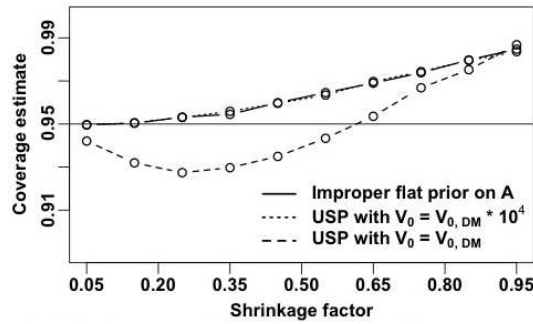


Figure 1: A partial result of our numerical illustration in Section 4.1 shows that there exists a choice for the shape parameter of the USP that meets the spirit of 95% confidence level better than an existent choice in the literature, e.g., $V_{0,DM}$ denotes a harmonic mean of $\{V_1, V_2, \dots, V_k\}$ suggested by DuMouchel (1994). See Section 4.1 for details.

(Lebesgue) on mp -dimensional space. The joint posterior density in (5) is proper when $k > p + m + 1$; see Theorem 2 of [Everson and Morris \(2000\)](#).

Since a rejection sampling for a multivariate Normal-Normal model ([Everson and Morris, 2000](#)) is not available for various choices for the shape parameter of the USP, we use a Metropolis-Hasting within Gibbs sampler ([Tierney, 1994](#)). This sampler iteratively samples the following conditional posterior distributions of the unknown parameters derived from (5); for $j = 1, 2, \dots, k$,

$$\boldsymbol{\theta}_j \mid \mathbf{A}, \boldsymbol{\beta}, \mathbf{y} \sim N_p \left((I_p - \mathbf{B}_j) \mathbf{y}_j + \mathbf{B}_j \mathbf{X}_j^\top \boldsymbol{\beta}, (I_p - \mathbf{B}_j) \mathbf{V}_j \right), \quad (6)$$

$$\boldsymbol{\beta} \mid \mathbf{A}, \boldsymbol{\theta}, \mathbf{y} \sim N_{mp}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \quad (7)$$

$$\pi_c(\mathbf{A} \mid \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{y}) \propto \left[\prod_{j=1}^k N_p(\boldsymbol{\theta}_j \mid \mathbf{X}_j^\top \boldsymbol{\beta}, \mathbf{A}) \right] \times |\mathbf{V}_0 + \mathbf{A}|^{-p-1} \times I_{\{|\mathbf{A}| > 0\}}, \quad (8)$$

where $\boldsymbol{\mu}_\beta = (\sum_{j=1}^k \mathbf{X}_j^\top \mathbf{A}^{-1} \mathbf{X}_j)^{-1} (\sum_{j=1}^k \mathbf{X}_j^\top \mathbf{A}^{-1} \boldsymbol{\theta}_j)$, $\boldsymbol{\Sigma}_\beta = (\sum_{j=1}^k \mathbf{X}_j^\top \mathbf{A}^{-1} \mathbf{X}_j)^{-1}$, and the notation $N_p(w \mid \mu, \Sigma)$ used in (8) is a p -dimensional Gaussian density of w with mean μ and covariance matrix Σ . At each iteration, we draw the unknown parameters in a sequence of (6), (7), and (8). It is easy to sample the multivariate Gaussian conditional posterior distributions in (6) and (7). However, the conditional posterior distribution of \mathbf{A} in (8) is not a standard family distribution that we can directly sample, and thus we use a Metropolis-Hastings algorithm to sample (8) within the Gibbs sampler. (If we used an improper flat prior distribution on \mathbf{A} , the conditional posterior distribution of \mathbf{A} in (8) would become an inverse Wishart($k - p - 1, \mathbf{S}$) distribution with degrees of freedom $k - p - 1$ and scale matrix $\mathbf{S} \equiv \sum_{j=1}^k (\boldsymbol{\theta}_j - \mathbf{X}_j^\top \boldsymbol{\beta})(\boldsymbol{\theta}_j - \mathbf{X}_j^\top \boldsymbol{\beta})^\top$.)

In a univariate case ($p = 1$), we draw $\log(A^*)$ from a Gaussian proposal distribution $N_1(\log(A^{(i-1)}) \mid \sigma^2)$ at iteration i , where the proposal scale σ is the tuning parameter. We set $A^{(i)}$ to A^* with probability

$$\min \left[1, \frac{\pi_c(A^* \mid \boldsymbol{\beta}^{(i)}, \boldsymbol{\theta}^{(i)}, \mathbf{y})}{\pi_c(A^{(i-1)} \mid \boldsymbol{\beta}^{(i)}, \boldsymbol{\theta}^{(i)}, \mathbf{y})} \times \frac{A^*}{A^{(i-1)}} \right] \quad (9)$$

and set $A^{(i)}$ to $A^{(i-1)}$ otherwise. The last term in (9), $A^*/A^{(i-1)}$, is the Hastings ratio for the update of A on a logarithmic scale. In a multivariate case, we set a proposal distribution of \mathbf{A} to an inverse Wishart distribution to avoid a matrix logarithm. At iteration i , we draw \mathbf{A}^* from the inverse Wishart($\nu, (\nu + p + 1) \times \mathbf{A}^{(i-1)}$) distribution with degrees of freedom ν and scale matrix $(\nu + p + 1) \times \mathbf{A}^{(i-1)}$; the mode of this inverse Wishart distribution is $\mathbf{A}^{(i-1)}$. We set $\mathbf{A}^{(i)}$ to \mathbf{A}^* with probability

$$\min \left[1, \frac{\pi_c(\mathbf{A}^* \mid \boldsymbol{\beta}^{(i)}, \boldsymbol{\theta}^{(i)}, \mathbf{y})}{\pi_c(\mathbf{A}^{(i-1)} \mid \boldsymbol{\beta}^{(i)}, \boldsymbol{\theta}^{(i)}, \mathbf{y})} \times \frac{\text{inv-W}(\mathbf{A}^{(i-1)} \mid \nu, (\nu + p + 1) \times \mathbf{A}^*)}{\text{inv-W}(\mathbf{A}^* \mid \nu, (\nu + p + 1) \times \mathbf{A}^{(i-1)})} \right],$$

where the notation $\text{inv-W}(\mathbf{A} \mid \nu, (\nu + p + 1) \times \Sigma)$ denotes the density function of the inverse Wishart($\nu, (\nu + p + 1) \times \Sigma$) distribution evaluated at \mathbf{A} . We treat the degrees of freedom ν as a tuning parameter for a reasonable acceptance rate.

In our numerical illustrations, we run the Metropolis-Hastings within Gibbs sampler for 42000 iterations, discarding the first 2000 iterations as burn-in. We then choose every other iteration of the Markov chain of length 40000, resulting in 20000 posterior samples of each parameter.

3 Multivariate frequency coverage evaluation

To check which choice for the shape parameter of the USP produces Bayesian interval estimates for random effects that meet their nominal confidence levels better, we conduct a repeated sampling coverage evaluation. Here we extend a univariate repeated sampling coverage evaluation specified in [Tak et al. \(2017\)](#) to a multivariate one. This evaluation is based on an assumption that our two-level Normal-Normal model is a true data generation process. We first generate n_{sim} sets of k random effects, $\{(\boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, \dots, \boldsymbol{\theta}_k^{(i)}), i = 1, 2, \dots, n_{\text{sim}}\}$, from the conjugate Gaussian prior distribution in (2) given specific values of \mathbf{A} and $\boldsymbol{\beta}$. Let us denote these generative values of \mathbf{A} and $\boldsymbol{\beta}$ by \mathbf{A}_{gen} and $\boldsymbol{\beta}_{\text{gen}}$, respectively. Given these random effects, we generate n_{sim} sets of k unbiased estimates $\{(\mathbf{y}_1^{(i)}, \mathbf{y}_2^{(i)}, \dots, \mathbf{y}_k^{(i)}), i = 1, 2, \dots, n_{\text{sim}}\}$ from the Gaussian distribution in (1).

We fit a Normal-Normal model on each simulation i , drawing 20000 posterior samples of $\boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, \dots, \boldsymbol{\theta}_k^{(i)}, \mathbf{A}$, and $\boldsymbol{\beta}$ using the Metropolis-Hasting within Gibbs sampler. For the l th component of $\boldsymbol{\theta}_j^{(i)}$, i.e., $\theta_{lj}^{(i)}$ for $l = 1, 2, \dots, p$, we calculate its 95% posterior interval by 0.025 and 0.975 quantiles of its 20,000 posterior samples that we denote by $(\theta_{lj,\text{low}}^{(i)}, \theta_{lj,\text{upp}}^{(i)})$. We define a coverage indicator of random effect j on the i -th simulation as

$$I(\boldsymbol{\theta}_j^{(i)}) = \prod_{l=1}^p I(\theta_{lj}^{(i)}), \quad (10)$$

where $I(\theta_{lj}^{(i)})$ on the right hand side represents another indicator function that takes the value one if the generative true value $\theta_{lj}^{(i)}$ lies within its 95% interval estimate $(\theta_{lj,\text{low}}^{(i)}, \theta_{lj,\text{upp}}^{(i)})$ and zero otherwise. Thus the coverage indicator $I(\boldsymbol{\theta}_j^{(i)})$ in (10) is equal to the value one if all the p components of the generative values $\boldsymbol{\theta}_j^{(i)}$ are within their 95% posterior intervals and zero otherwise.

We assume that there exists an unknown coverage probability of random effect j denoted by $C_{\mathbf{A}_{\text{gen}}, \boldsymbol{\beta}_{\text{gen}}}(\boldsymbol{\theta}_j)$ depending on the values of \mathbf{A} and $\boldsymbol{\beta}$ that generate random effects and simulated data sets. The coverage indicator $I(\boldsymbol{\theta}_j^{(i)})$ in (10) is assumed to follow an independent and identically distributed Bernoulli distribution given the unknown coverage rate $C_{\mathbf{A}_{\text{gen}}, \boldsymbol{\beta}_{\text{gen}}}(\boldsymbol{\theta}_j)$. The mean of the coverage indicators averaged over n simulations, i.e., $\bar{I}(\boldsymbol{\theta}_j) = \sum_{i=1}^{n_{\text{sim}}} I(\boldsymbol{\theta}_j^{(i)}) / n_{\text{sim}}$, is a simple unbiased coverage estimator for $C_{\mathbf{A}_{\text{gen}}, \boldsymbol{\beta}_{\text{gen}}}(\boldsymbol{\theta}_j)$. We define a more efficient Rao-Blackwellized (RB) unbiased coverage

estimator as

$$\begin{aligned}\bar{I}^{RB}(\boldsymbol{\theta}_j) &= \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} E\left(I(\boldsymbol{\theta}_j^{(i)}) \mid \mathbf{A}_{\text{gen}}, \boldsymbol{\beta}_{\text{gen}}, \mathbf{y}_j^{(i)}\right) \\ &= \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} P\left(\theta_{1j}^{(i)} \in \left(\theta_{1j,\text{low}}^{(i)}, \theta_{1j,\text{upp}}^{(i)}\right), \dots, \right. \\ &\quad \left. \theta_{pj}^{(i)} \in \left(\theta_{pj,\text{low}}^{(i)}, \theta_{pj,\text{upp}}^{(i)}\right) \mid \mathbf{A}_{\text{gen}}, \boldsymbol{\beta}_{\text{gen}}, \mathbf{y}_j^{(i)}\right).\end{aligned}\tag{11}$$

The expectation of $\bar{I}^{RB}(\boldsymbol{\theta}_j)$ given \mathbf{A}_{gen} and $\boldsymbol{\beta}_{\text{gen}}$ is $C_{\mathbf{A}_{\text{gen}}, \boldsymbol{\beta}_{\text{gen}}}(\boldsymbol{\theta}_j)$. We use the cumulative distribution of the conditional posterior distribution of random effect j in (3) to compute the conditional posterior probabilities in (11). The variance of $\bar{I}^{RB}(\boldsymbol{\theta}_j)$ does not exceed that of $\bar{I}(\boldsymbol{\theta}_j)$ because of the Rao-Blackwellization (Rao, 1945; Blackwell, 1947), and its unbiased estimate is

$$\widehat{\text{Var}}(\bar{I}^{RB}(\boldsymbol{\theta}_j)) = \frac{1}{n_{\text{sim}}(n_{\text{sim}} - 1)} \sum_{i=1}^{n_{\text{sim}}} \left(E\left(I(\boldsymbol{\theta}_j^{(i)}) \mid \mathbf{A}_{\text{gen}}, \boldsymbol{\beta}_{\text{gen}}, \mathbf{y}_j^{(i)}\right) - \bar{I}^{RB}(\boldsymbol{\theta}_j)\right)^2.$$

To summarize the results of the frequency coverage evaluation, we report the following overall coverage estimate and its variance estimate:

$$\bar{I}^{RB} = \frac{1}{k} \sum_{j=1}^k \bar{I}^{RB}(\boldsymbol{\theta}_j) \quad \text{and} \quad \widehat{\text{Var}}(\bar{I}^{RB}) = \frac{1}{k^2} \sum_{j=1}^k \widehat{\text{Var}}(\bar{I}^{RB}(\boldsymbol{\theta}_j)).\tag{12}$$

4 Numerical illustrations

Using the frequency coverage evaluation, we show which choice for the shape parameter produces Bayesian interval estimates for random effects that meet their nominal confidence levels better. The first numerical illustration is to investigate the shape parameter V_0 of the univariate USP ($p = 1$) in a simple setting and the second one is to see whether this univariate case can be consistently extended to a multivariate USP ($p = 2$).

4.1 Eight school data

To test the effect of coaching on students' SAT scores, the Education Testing Service conducted randomized experiments in eight schools (Rubin, 1981). The data are composed of the estimated coaching effects on SAT scores (y_j) and standard errors of the eight schools ($j = 1, 2, \dots, 8$). The data appear in Table 1. We assume each school's coaching effect approximately follows a Gaussian sampling distribution and its sampling variance (V_j) is approximately known, considering a large number of students tested in each school (Tak et al., 2017). Since no covariate information is available, we assume $\mathbf{X}_j^\top \boldsymbol{\beta} = \beta_1$, leading to an exchangeable Gaussian prior distribution in (2).

We begin the frequency coverage evaluation by choosing the generative values, \mathbf{A}_{gen} and $\boldsymbol{\beta}_{1,\text{gen}}$, in a way to favor the existent choice for V_0 , e.g., the harmonic mean of

Table 1: The data are composed of the effect of coaching on students' SAT scores (y_j) and its standard error ($V_j^{1/2}$) for each of the eight schools ($j = 1, 2, \dots, 8$).

j	1	2	3	4	5	6	7	8
y_j	28	8	-3	7	-1	1	18	12
$V_j^{1/2}$	15	10	16	11	9	11	10	18

$\{V_1, V_2, \dots, V_k\}$, denoted by $V_{0,DM}$ (DuMouchel, 1994). To see the pattern of coverage rates according to A_{gen} , we choose ten values of A_{gen} by converting ten values of the shrinkage factor $B_0 \in \{0.05, 0.15, \dots, 0.95\}$, using the relationship $A = (1 - B_0)V_0/B_0$. The resulting ten values of A_{gen} with V_0 set to $V_{0,DM} = 132.6$ are $\{2520.2, 751.7, 397.9, 246.3, 162.1, 108.5, 71.4, 44.2, 23.4, 7.0\}$ and we denote these by $A_{\text{gen},1}, A_{\text{gen},2}, \dots, A_{\text{gen},10}$. To set $\beta_{1,\text{gen}}$, we fit a Normal-Normal model using a USP with $V_{0,DM}$, and set $\beta_{1,\text{gen}}$ to 7.95, the posterior mean of β_1 based on its 100,000 posterior samples. Finally we have ten pairs of the generative values, i.e., $(A_{\text{gen},1}, \beta_{1,\text{gen}}), (A_{\text{gen},2}, \beta_{1,\text{gen}}), \dots, (A_{\text{gen},10}, \beta_{1,\text{gen}})$.

We compare the frequency coverage properties of six different prior distributions on A ; the improper flat prior distribution on A and the USPs on A with five different shape parameters, i.e., $V_0 = \delta \times V_{0,DM}$ for $\delta \in \{10^0, 10^1, 10^2, 10^3, 10^4\}$. For each prior distribution on A , we conduct ten frequency coverage evaluations using ten different pairs of the generative values, i.e., we conduct the frequency coverage evaluation 60 times in total for the combinations of ten pairs of the generative values and six different prior distributions on A . For each frequency coverage evaluation, we generate 1000 ($= n_{\text{sim}}$) mock data sets. On each of the generated data sets, we fit a Normal-Normal model equipped with the corresponding prior distribution on A and an improper flat prior distribution on β_1 , drawing 20000 posterior samples of $\theta_1, \theta_2, \dots, \theta_8, A$, and β_1 . To initialize the Metropolis-Hastings within Gibbs sampler, we set $\theta_j^{(0)} = y_j$ for all j , $\beta_1^{(0)} = \bar{y}$, and $A^{(0)} = V_{0,DM}$. We set the scale of the Gaussian proposal distribution for $\log(A)$ to two, i.e., $\sigma = 2$, resulting in an average acceptance rate equal to 0.326 across all the simulations.

Using the 20000 posterior samples of each random effect, whose effective sample size is 8872 on average across all the simulations, we calculate 95% posterior intervals for random effects, i.e., $(\theta_{j,\text{low}}^{(i)}, \theta_{j,\text{upp}}^{(i)})$ for $j = 1, 2, \dots, 8$ and $i = 1, 2, \dots, 20000$. (Subscript l used in (10) is suppressed for simplicity because $p = 1$.) Then, we calculate the RB coverage rate estimates, $\bar{I}^{RB}(\theta_j)$ in (11). We summarize the overall coverage estimates and their standard errors defined in (12) for three different prior distributions on A in Table 2 and display the overall coverage estimates for all 60 cases in Figure 2. The standard errors are too small to be displayed.

In the left panel of Figure 2, the overall coverage estimates based on the USP with the existent choice $V_{0,DM}$, denoted by the dashed line, do not achieve the 95% confidence level over a wide range of the generative values of $B_{0,i} = V_{0,DM}/(V_{0,DM} + A_{\text{gen},i})$ between 0.05 and 0.55. However, as the USP on A approaches the improper flat prior on A , i.e., as V_0 increases, the estimated coverage rate of the USP converges to that of the improper

Table 2: The estimated coverage rates and their standard errors (in parentheses) specified in (12). The confidence level is set to 0.95. Three different prior distributions on A appear in columns and ten different generative values of $B_{0,i} = V_{0,DM}/(V_{0,DM} + A_{gen,i})$ appear in rows. As the shape parameter of the USP increases, the coverage estimates based on the USP approach those based on the improper flat prior on A , conservatively achieving the 95% confidence level.

	USP with $V_0 = V_{0,DM}$	USP with $V_0 = 10^4 \times V_{0,DM}$	Improper flat on A
$B_{0,1} = 0.05$	0.943 (0.0002)	0.950 (0.0001)	0.950 (0.0001)
$B_{0,2} = 0.15$	0.931 (0.0005)	0.951 (0.0002)	0.950 (0.0002)
$B_{0,3} = 0.25$	0.926 (0.0005)	0.952 (0.0003)	0.953 (0.0002)
$B_{0,4} = 0.35$	0.930 (0.0005)	0.955 (0.0003)	0.954 (0.0003)
$B_{0,5} = 0.45$	0.936 (0.0005)	0.959 (0.0003)	0.960 (0.0003)
$B_{0,6} = 0.55$	0.944 (0.0004)	0.963 (0.0004)	0.965 (0.0004)
$B_{0,7} = 0.65$	0.956 (0.0004)	0.969 (0.0005)	0.969 (0.0005)
$B_{0,8} = 0.75$	0.967 (0.0005)	0.974 (0.0007)	0.974 (0.0006)
$B_{0,9} = 0.85$	0.977 (0.0006)	0.980 (0.0007)	0.980 (0.0007)
$B_{0,10} = 0.95$	0.988 (0.0007)	0.983 (0.0010)	0.985 (0.0009)

flat prior on A , achieving the 95% confidence level better; see the right panel of Figure 2. This is because the USP, $(V_0 + A)^{-2}dA$, puts more mass at $A = 0$ *a priori* when V_0 is small. Since the resulting posterior inference can be affected strongly by the prior distribution when the data size is small, the smaller shape parameter V_0 in the USP

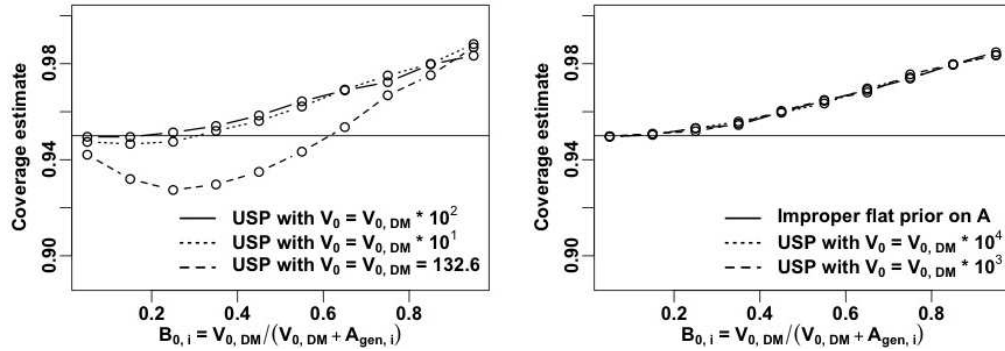


Figure 2: Estimated coverage rates obtained from the frequency coverage evaluations. The overall coverage estimates based on $V_{0,DM} = 132.6$, denoted by the dashed line on the left panel, do not meet the 95% confidence level over a wide range of the generative values of $B_{0,i}$ between 0.05 and 0.55. As the USP on A approaches the improper flat prior on A , i.e., as V_0 increases, the estimated coverage rates of the USP converge to those of the improper flat prior on A , meeting the 95% confidence level better.

results in a smaller posterior sample of A , a larger posterior sample of $B_j = V_j/(V_j + A)$, and thus shorter posterior intervals for θ_j as implied by the variance in (3). Consequently, these shorter posterior intervals produce smaller coverage estimates. The opposite occurs by flattening the USP on A with larger V_0 ; the resulting posterior intervals for random effects become wider, leading to more conservative posterior inference on random effects.

The right panel of Figure 2 shows that the estimated coverage rates do not change much when δ in $V_0 = \delta \times V_{0,DM}$ is greater than or equal to 10^3 . This indicates that a safe choice for V_0 to achieve good frequency coverage properties in the random effects estimation is to set a large enough value for V_0 compared to a typical value of V_j 's, e.g., $V_0 = 10^4 \times V_{0,DM}$ or $V_0 = 10^4 \times V_{0,E\&M}$, where $V_{0,E\&M}$ is the arithmetic mean of $\{V_1, V_2, \dots, V_k\}$ suggested by [Everson and Morris \(2000\)](#).

4.2 Twenty seven hospital profiling data

The twenty seven hospital profiling data are based on the interviews of 1869 patients from teaching hospitals and academic health centers ([Cleary et al., 1991](#)). The data are composed of the number of interviewees (n_j), the severity measure averaged over the health indices of interviewees (x_{2j}), the percentage of non-surgical problems (Y_{1j}), and that of surgical problems (Y_{2j}) experienced by patients in hospital j ($j = 1, 2, \dots, 27$). (We fit an intercept term by setting $x_{1j} = 1$ for all j .) The data are tabulated in Table 3. [Everson and Morris \(2000\)](#) used these data to show the effectiveness of their rejection sampling based on the improper flat prior on \mathbf{A} compared to the REML method. They

Table 3: The twenty seven hospital profiling data are composed of the percentage of non-surgical problems (Y_{1j}) experienced by patients in hospital j , that of surgical problems (Y_{2j}), the severity measure averaged over the health indices of interviewees (x_{2j}), and the number of interviewees (n_j). This table is reproduced from [Everson and Morris \(2000\)](#).

j	Y_{1j}	Y_{2j}	x_{2j}	n_j	j	Y_{1j}	Y_{2j}	x_{2j}	n_j
1	10.18	15.06	0.75	24	15	15.80	11.50	0.26	61
2	11.55	17.97	0.62	32	16	14.81	20.56	0.56	62
3	16.21	12.50	0.66	32	17	11.14	13.02	0.02	62
4	12.31	14.88	0.26	43	18	17.12	14.60	0.41	66
5	12.88	15.21	0.96	44	19	16.93	16.28	0.56	68
6	11.84	17.69	0.44	45	20	11.02	13.52	0.34	68
7	14.82	16.91	0.44	48	21	14.69	16.49	0.56	72
8	13.05	15.07	0.55	49	22	10.48	14.24	0.79	77
9	12.43	12.01	0.33	51	23	15.82	15.13	0.47	87
10	8.35	9.43	0.47	53	24	12.66	14.99	0.71	122
11	17.97	26.82	0.48	56	25	10.41	17.25	0.45	124
12	11.84	15.64	0.34	58	26	10.32	10.13	0.05	149
13	12.43	13.94	0.28	58	27	13.72	18.18	0.77	198
14	14.73	15.40	0.63	60					

estimated a common covariance matrix of the percentages of non-surgical and surgical problems from the entire date of the 1869 patients and set it to Σ as follows;

$$\Sigma = \begin{pmatrix} 148.87 & 140.43 \\ 140.43 & 490.60 \end{pmatrix}.$$

Using this Σ , they set the covariance matrix of hospital j to $V_j = \Sigma/n_j$. Following [Everson and Morris \(2000\)](#), we assume that the sampling distribution of (Y_{1j}, Y_{2j}) approximately follows a bivariate Gaussian distribution with known covariance matrix V_j , considering a reasonably large number of interviewees in each hospital.

We start the frequency coverage evaluation by setting A_{gen} and β_{gen} in such a way to favor the existent choice for V_0 , e.g., an arithmetic mean of $\{V_1, V_2, \dots, V_k\}$, denoted by $V_{0,E\&M}$ ([Everson and Morris, 2000](#)). We first fit a bivariate Normal-Normal model using a USP with $V_{0,E\&M}$ and a flat prior distribution on β and set β_{gen} to the posterior means of β based on its 100000 posterior samples. We set $A_{\text{gen},i} = \Sigma/u_i$ for $i = 1, 2, \dots, 10$, where u_i is the i th element of $(0.29, 0.63, 1.00, 1.45, 2.04, 2.87, 4.16, 6.47, 11.82, 38.50)$ so that the resulting ten values of $|B_{0,i}| \equiv |V_{0,E\&M}(V_{0,E\&M} + A_{\text{gen},i})^{-1}|$ are $(0.05, 0.15, \dots, 0.95)$. Finally, we obtain ten different combinations of the generative values, $(A_{\text{gen},1}, \beta_{\text{gen}}), \dots, (A_{\text{gen},10}, \beta_{\text{gen}})$.

We compare operating characteristics of six different prior distributions on A . We use five USPs with five different shape parameters, i.e., $V_0 = V_{0,E\&M}$ and $V_0 = \delta \times \text{diag}_2(V_{0,E\&M})$ for $w \in \{10^1, 10^2, 10^3, 10^4\}$, where the notation $\text{diag}_p(V)$ represents a p -dimensional diagonal matrix whose diagonal elements are the same as those of V . The sixth prior distribution on A is an improper flat prior distribution, i.e., $I_{\{|A|>0\}}dA$. For each of the 60 frequency coverage evaluations (ten pairs of the generative values and six different prior distributions on A), we generate 1000 ($= n_{\text{sim}}$) mock data sets. For each simulation, we fit a bivariate Normal-Normal model equipped with the corresponding prior distribution of A and $d\beta$, drawing 20000 posterior samples of θ , A , and β using the Metropolis-Hastings within Gibbs sampler. We set the initial value of θ_j to y_j for all j , of β to β_{gen} , and of A to $V_{0,E\&M}$. We set the degrees of freedom ν of the inverse Wishart proposal distribution for A to 40, leading to the average acceptance rate equal to 0.378 across all the simulations. The average effective sample size for each random effect is 6817 across all the simulations. We calculate 95% posterior intervals $(\theta_{lj,\text{low}}^{(i)}, \theta_{lj,\text{upp}}^{(i)})$ for $l = 1, 2$, $j = 1, 2, \dots, 27$ and $i = 1, 2, \dots, 20000$, and compute the RB coverage rate estimates, $\bar{I}^{RB}(\theta_j)$ in (11). Table 4 summarizes the overall coverage estimates, \bar{I}^{RB} in (12) and their standard errors for three different prior distributions. Figure 3 shows the overall coverage estimates for all 60 cases. The standard errors are too small to be displayed.

The results displayed in Figure 3 are similar to those of the univariate case ($p = 1$) in Section 4.1. In the left panel of Figure 3, the coverage estimates based on the existent choice $V_{0,E\&M}$, denoted by the dashed line, do not achieve the 95% confidence level over the range of the generative values of $|B_{0,i}| = |V_{0,E\&M}(V_{0,E\&M} + A_{\text{gen},i})^{-1}|$ between 0.05 and 0.55. However, as the USP approaches the improper flat prior on A , i.e., as δ in $V_0 = \delta \times \text{diag}_2(V_{0,E\&M})$ increases, the estimated coverage rate of the USP approaches that of the improper flat prior on A as shown in the right panel. This is because the

Table 4: The coverage estimates and their standard errors (in parentheses) specified in (12). The confidence level is set to 0.95. Three different prior distributions on A appear in columns and ten determinants of the generative values of $B_{0,i} = V_{0,E\&M} / (V_{0,E\&M} + A_{\text{gen},i})$ appear in rows. As δ in $V_0 = \delta \times \text{diag}_2(V_{0,E\&M})$ increases, the coverage estimates based on the USP approach those based on the improper flat prior on A , conservatively achieving the 95% confidence level. The notation $\text{diag}_p(V)$ represents a p -dimensional diagonal matrix whose diagonal elements are the same as those of V .

	USP with $V_0 = V_{0,E\&M}$	USP with $V_0 = 10^4 \times \text{diag}_2(V_{0,E\&M})$	Improper flat on A
$ B_{0,1} = 0.05$	0.941 (0.0002)	0.950 (0.0001)	0.949 (0.0001)
$ B_{0,2} = 0.15$	0.933 (0.0003)	0.950 (0.0002)	0.950 (0.0002)
$ B_{0,3} = 0.25$	0.935 (0.0003)	0.952 (0.0002)	0.952 (0.0002)
$ B_{0,4} = 0.35$	0.937 (0.0003)	0.956 (0.0002)	0.956 (0.0002)
$ B_{0,5} = 0.45$	0.945 (0.0003)	0.962 (0.0002)	0.961 (0.0002)
$ B_{0,6} = 0.55$	0.955 (0.0003)	0.968 (0.0002)	0.968 (0.0002)
$ B_{0,7} = 0.65$	0.967 (0.0002)	0.974 (0.0002)	0.975 (0.0003)
$ B_{0,8} = 0.75$	0.976 (0.0003)	0.981 (0.0003)	0.982 (0.0002)
$ B_{0,9} = 0.85$	0.987 (0.0003)	0.987 (0.0003)	0.988 (0.0003)
$ B_{0,10} = 0.95$	0.994 (0.0003)	0.992 (0.0004)	0.993 (0.0003)

large value of δ flattens the USP on A and thus the coverage estimates based on the

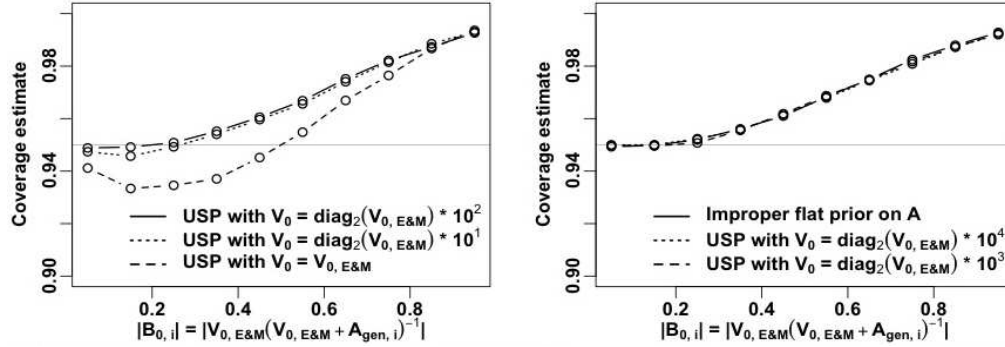


Figure 3: Estimated coverage rates from the frequency coverage evaluation. The coverage estimates based on $V_{0,E\&M}$, denoted by the dashed line on the left panel exhibit under-coverages over a wide range of the generative values of $|B_{0,i}| = |V_{0,E\&M}(V_{0,E\&M} + A_{\text{gen},i})^{-1}|$ between 0.05 and 0.55. As the USP on A approaches the improper flat prior distribution on A , i.e., as δ in $V_0 = \delta \times \text{diag}_2(V_{0,E\&M})$ increases, the estimated coverage rate of the USP converges to that of the improper flat prior distribution on A , meeting the 95% confidence level better. The notation $\text{diag}_p(V)$ represents a p -dimensional diagonal matrix whose diagonal elements are the same as those of V .

USP approach those based on the improper flat prior on \mathbf{A} .

The right panel of Figure 3 shows that the estimated coverage rates do not change much when δ in $\mathbf{V}_0 = \delta \times \text{diag}_2(\mathbf{V}_{0,\text{E\&M}})$ is greater than or equal to 10^3 . This indicates that it is safe to set a large enough value of δ that leads to a large value of $|\mathbf{V}_0|$ compared to a typical value of $|\mathbf{V}_j|$'s to achieve better frequency coverage properties in the random effects estimation, e.g., $\mathbf{V}_0 = 10^4 \times \text{diag}_2(\mathbf{V}_{0,\text{E\&M}})$ or $\mathbf{V}_0 = 10^4 \times \text{diag}_2(\mathbf{V}_{0,\text{DM}})$.

5 Concluding Remarks

A uniform shrinkage prior (USP) on the unknown variance component of a random-effects model is known to produce good frequency properties (Christiansen and Morris, 1997; Daniels, 1999; Natarajan and Kass, 2000). However, it has been neglected whether the USP can maintain such good frequency properties regardless of its shape parameter. We implement frequency coverage evaluations to see which choice for the shape parameter of the USP enables the posterior intervals for random effects to meet their nominal confidence levels better. In our numerical illustrations, the shape parameter that flattens the USP on \mathbf{A} achieves better frequency coverage properties in estimating random effects than the existent choices suggested in the literature.

There are several opportunities to build upon our work. Implementing the frequency coverage evaluation is computationally expensive because it fits a model on every simulated data set. For example, it takes almost a week to complete one frequency coverage evaluation that fits the model on a thousand simulated data sets in Section 4.2. Thus, it is desirable to develop an independent sampler for the Normal-Normal model with the USP instead of a Markov chain Monte Carlo method used in this article. Also, the frequency coverage properties for the USP according to the shape parameter have not been tested in different hierarchical models, e.g., Poisson-Gamma or Beta-Binomial models, though they may produce similar results asymptotically. Another avenue for further improvement is to check high-dimensional behavior of the USP according to the shape parameter. We leave these for our future research.

References

- Albert, J. H. (1988). "Computational Methods using a Bayesian Hierarchical Generalized Linear Model." *Journal of the American Statistical Association*, 83(404): 1037–1044.
- Blackwell, D. (1947). "Conditional Expectation and Unbiased Sequential Estimation." *The Annals of Mathematical Statistics*, 18(1): 105–110.
- Chen, H.-C. and Wehrly, T. E. (2016). "Approximate Uniform Shrinkage Prior for a Multivariate Generalized Linear Mixed Model." *Journal of Multivariate Analysis*, 145: 148–161.
- Christiansen, C. L. and Morris, C. N. (1997). "Hierarchical Poisson Regression Modeling." *Journal of the American Statistical Association*, 92(438): 618–632.

- Cleary, P. D., Edgman-Levitan, S., Roberts, M., Moloney, T. W., McMullen, W., Walker, J. D., and Delbanco, T. L. (1991). “Patients Evaluate Their Hospital Care: A National Survey.” *Health Affairs*, 10(4): 254–267.
- Daniels, M. J. (1999). “A Prior for the Variance in Hierarchical Models.” *Canadian Journal of Statistics*, 27(3): 567–578.
- DuMouchel, W. (1994). “Hierarchical Bayes Linear Models for Meta-Analysis.” *National Institute of Statistical Sciences Technical Report No. 27*.
- Everson, P. J. and Morris, C. N. (2000). “Inference for Multivariate Normal Hierarchical Models.” *Journal of the Royal Statistical Society B*, 62(2): 399–412.
- Natarajan, R. and Kass, R. E. (2000). “Reference Bayesian Methods for Generalized Linear Mixed Models.” *Journal of the American Statistical Association*, 95(449): 227–237.
- Rao, C. R. (1945). “Information and Accuracy Attainable in the Estimation of Statistical Parameters.” *Bulletin of the Calcutta Mathematical Society*, 37(3): 81–91.
- Rubin, D. B. (1981). “Estimation in Parallel Randomized Experiments.” *Journal of Educational Statistics*, 6(4): 377–401.
- Strawderman, W. E. (1971). “Proper Bayes Minimax Estimators of the Multivariate Normal Mean.” *The Annals of Mathematical Statistics*, 42(1): 385–388.
- Tak, H., Kelly, J., and Morris, C. N. (2017). “Rgbr: An R Package for Gaussian, Poisson, and Binomial Random Effects Models with Frequency Coverage Evaluations.” *Journal of Statistical Software*.
- Tierney, L. (1994). “Markov Chains for Exploring Posterior Distributions.” *The Annals of Statistics*, 22(4): 1701–1728.

Acknowledgments

The author acknowledges partial support from the NSF under Grant DMS 1127914 to the Statistical and Applied Mathematical Sciences Institute and thanks Carl N. Morris for very helpful discussions and Steven R. Finch for proofreading..